Demonstration of an effective Flexible Mask Optimization (FMO) flow

Charlotte Beylier[A], Nicolas Martin[B], Vincent Farys[A], Franck Foussadier[A], Emek Yesilada[A], Frederic Robert[A], Stanislas Baron[B], Russell Dover[B], Hua-yu Liu[B]

[A] STMicroelectronics, 850, rue Jean Monnet, FR-38921 Crolles cedex, France

[B] ASML Brion Technologies, 4211 Burton Dr., Santa Clara, CA, 95054, USA.

ABSTRACT

The 2x nm generation of advanced designs presents a major lithography challenge to achieve adequate correction due to the very low k1 values. The burden thus falls on resolution enhancement techniques (RET) in order to be able to achieve enough image contrast, with much of this falling to computational lithography. Advanced mask correction techniques can be computationally expensive. This paper presents a methodology that enables advanced mask quality with the cost of much simpler methods. Brion Technologies has developed a product called Flexible Mask Optimization (FMO) which identifies hotspots, applies an advanced technique to improve them, performs model based boundary healing to reinsert the repaired hotspot cleanly (without introducing new hotspots), and then performs a final verification. STMicroelectronics has partnered with Brion to evaluate and prove out the capability and performance of this approach.

The results shown demonstrate improved performance on 2x nm node complex 2D hole layers using a hybrid approach of rule based sub resolution assist features (RB-SRAF) and model based SRAF (MB-SRAF). The effective outcome is to achieve MB-SRAF levels of quality but at only a slightly higher computational cost than a quick, cheap rule based approach.

Keywords: FMO, MB-SRAF, PWOPC, LMC.

1. INTRODUCTION

Shrink of design dimensions in the semiconductor industry forces resolution enhancement techniques (RET) to become more and more performing in order to achieve physical manufacturability. Particularly for 2x nm generations, which have very low k1 values, every new technique is welcomed to help improving resolution. To go through this challenge keeping production constrains in mind, a compromise has to be found between mask optimization performance and costs [1].

Sub resolution assist features (SRAF) have become essential to get enough contrast, and are largely used in production with a rule based (RB) approach to attain sufficient process window (PW). Even if this approach shows very good results in terms of resolution and cost, some hotspot configuration may happen and reveal improvement opportunities. To push forward the SRAF insertion quality, a model based (MB) approach is a candidate to bridge the gap of hotspots resolution.

In this paper, ST Microelectronics has evaluated a new methodology based on FMO, developed by ASML Brion Technologies. This flow uses a MB-SRAF insertion method to apply enhancement on those hotspot configurations using a rule based SRAF (RB-SRAF) technique. This method unites the MB accuracy of SRAF generation with a flexible and fast flow. We will discuss about the gain of MB-SRAF compared to RB-SRAF, and then describe the entire flow with its results on a concrete case.

2. TACHYON FMO CONCEPT OVERVIEW

Tachyon FMO option provides a highly flexible capability to enhance the performance and cost effectiveness of Tachyon OPC+ and Tachyon LMC. Tachyon FMO enables the user to apply advanced OPC techniques to deliver improved performance in localized areas, only where necessary, therefore achieving optimum process performance with the least tape-out and mask cost. Tachyon FMO makes it possible to use and take advantage of multiple mask correction approaches with high quality boundary healing between the different correction areas. Thus, different areas within a design can have different correction techniques applied without inducing hotspots in the boundary regions between them. Critical to enabling Tachyon FMO is the detection and manipulation of hotspots [2] as well as the ability to cleanly

Optical Microlithography XXV, edited by Will Conley, Proc. of SPIE Vol. 8326, 832616

reinsert the corrected hotspots back into the full chip design without introducing new defects due to proximity effects of neighboring features.

Figure 1 represents an enhancement area where a hotspot has been detected at necking condition with a dose of -4 % and a focus of + 60 nm. In this example, a different SRAF solution is applied in the enhancement area in order to replace the current RB-SRAF. This repair leads to a higher PW at necking condition in the case of contact layer (Figure 2).

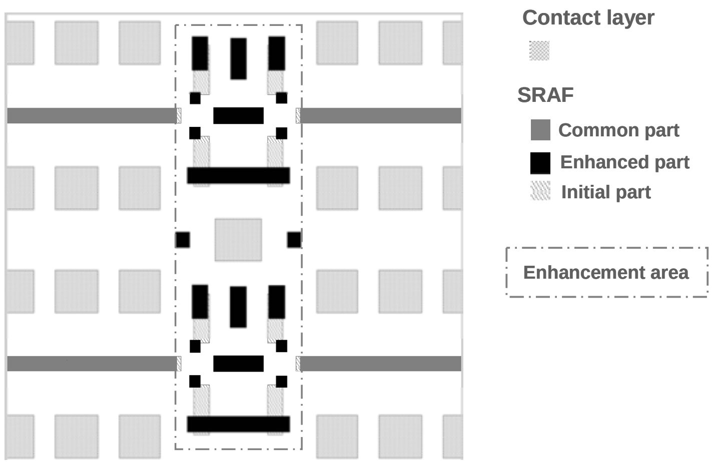


Figure 1. Definition of SRAF enhancement area

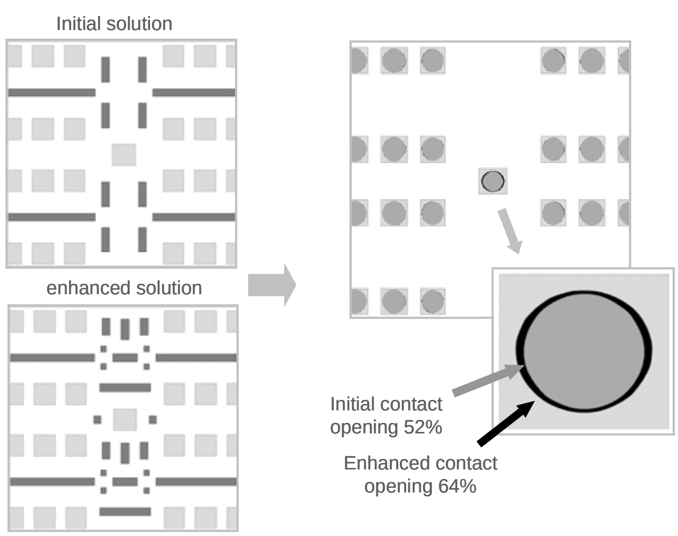


Figure 2. Example of FMO enhancement on contact layer at necking condition

2xnm designs require more advanced techniques to achieve pattern fidelity in lithography compared to any prior technology generation. Techniques such as MB-SRAF and 3D mask modeling (M3D) are computationally expensive and can add complexity and costs to mask making. Therefore any technique which increases effective throughput (at the same or lower cost) and reduces mask complexity has a strong return on investment for the user.

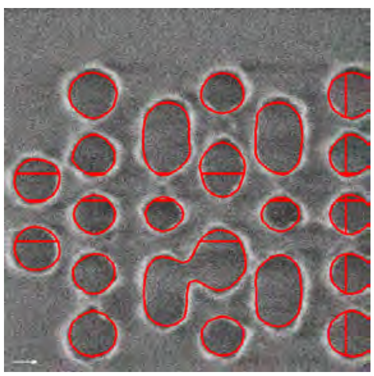
3. METRIC TO DETECT WORST CONFIGURATION

One of the main challenges for 2xnm generation node is the ability to properly monitor the lithographic PW. Advanced nodes are close to scanner PW margin, around 80 nm of depth of focus (DOF) at 5% of exposure latitude, due to the fact

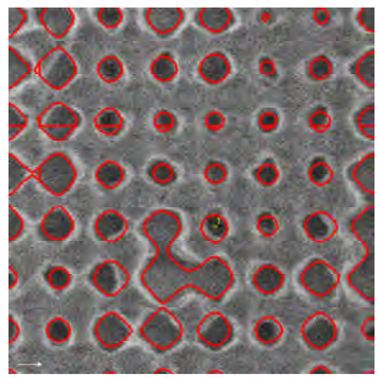
that optical resolution limit is reached [3]. Moreover, even the use of a smart solution in the SRAF placement permits us to have only a slight improvement in term of DOF (~ 10 nm DOF improvement). As such it is quite hard to make an experimental comparison between two SRAF solutions simply looking at CD SEM measurement on resist. Indeed, the CD difference between these two techniques is in the order of SEM uncertainty magnitude or scanner variability effects.

The ideal approach, to compare two SRAF techniques, is to use electrical or yield analysis in order to go beyond metrology limitation. But this approach is time and cost consuming. An alternate solution is to stay at the lithographic level and make full chip comparison directly on simulated contours to get a large number of statistical data.

This approach implies that the resist model has to be accurately calibrated through the lithographic PW. Figure 3 displays SEM images overlapped with simulated contours for two different structures. Figure 3.a shows 28 nm node SRAM cell at contact level with an overexposed dose. We clearly see that the simulated contour in red fits the SEM image perfectly. Figure 3.b represents the application of the model on unsupported structures which is a contact array with post OPC CDs varying within a range of 2nm. Again, there is a good correlation between SEM measurements and simulated contours. Thus, simulated contours can be followed to perform statistical analysis.



a) SRAM cells



b) Dense array structure

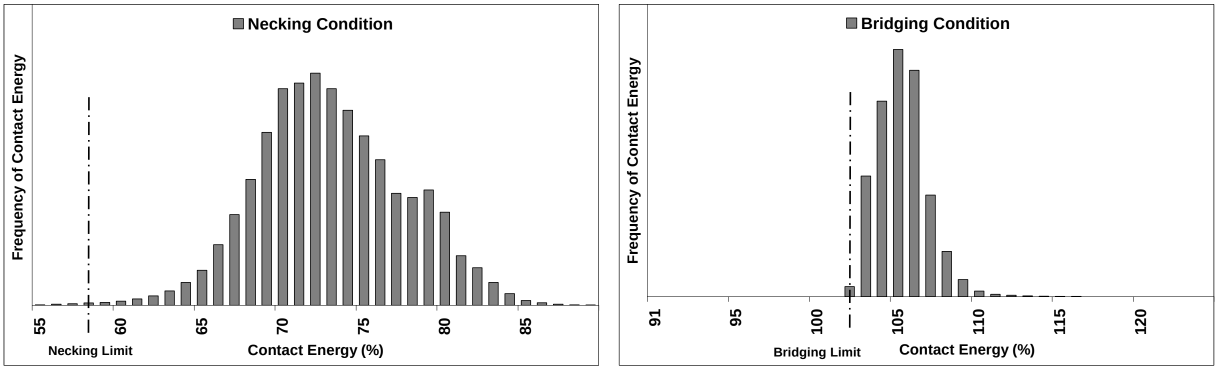
Figure 3. Overlap of SEM image with simulated contour (red line) for contact 28 nm node

We propose to use the '*Contact Energy*' metric to make the comparison between different SRAF techniques. The *Contact Energy* represents the ratio of the area of a simulated contour at a given PW condition to the area of the curved target (see Equation 1).

$$Contact\quad Energy(\%) = \frac{Area\quad Contour_{PW}}{Area\quad Contour_{\mathrm{Re}\,f}}.100 \tag{1}$$

The Figure 4 illustrates the contact energy histogram for two PW conditions. Figure 4.a shows the dispersion for a dose of -4% and a defocus of +60nm. For that condition, the contact energy is below 100 % which means that there is a risk of resist necking which leads to risk of missing contact after etch. On the Figure 4.b contact energy distribution is reported for a dose of +4%, a focus of 0nm and a bias of 0.5nm which represents a risk of contact bridging. For the rest of the paper, only the necking condition is discussed which is the most critical configuration for contact. Bridging risk is deemed less critical due to fewer occurrences on a full chip and smaller contact energy variation in process window compared to necking.

Based on these charts, we can also determine a threshold linked with the best PW performance that we can achieve with the lithographic process. In the necking condition case, this threshold represents the smallest contact energy reached for contact array that has a pitch at the limit of SRAF insertion. Assuming this threshold, it means that all contact cases which have contact energy below this limit can be enhanced by adjusting the SRAF placement. For our application, we consider that all contacts which have contact energy below this limit are considered as a hotspot, and propose to develop a flow to fix all these limited cases by bringing them above the limit.



a) Necking condition: dose -4%, focus = + 60 nm b) Bridging condition: +4%, 0 nm, bias 0.5nm

Figure 4. Contact energy histogram

4. APPLICATION ON 28 NM NODE CONTACT LAYER

4.1 Full chip RB-SRAF versus full chip MB-SRAF

SRAF insertion is one of the inevitable mask optimization techniques of hole levels, for enlarging process window. In this paragraph, a comparison is made between the RB-SRAF insertion used in production and the MB-SRAF insertion on 28nm contact layer.

By definition, the RB-SRAF insertion refers to standard rules, which define the allowed placements between SRAF to main pattern, SRAF to SRAF and size of the polygons (see figure 5). The SRAF printability is the limiting factor of the rules definition. Nevertheless, in some hotspot configurations the rule based placement may not be efficient enough because those rules may not be adapted to every situation. This limit can be overcome with a MB-SRAF approach.

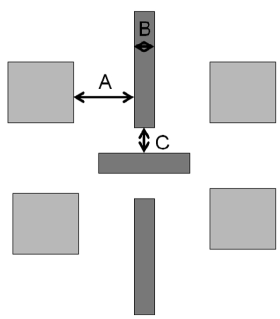
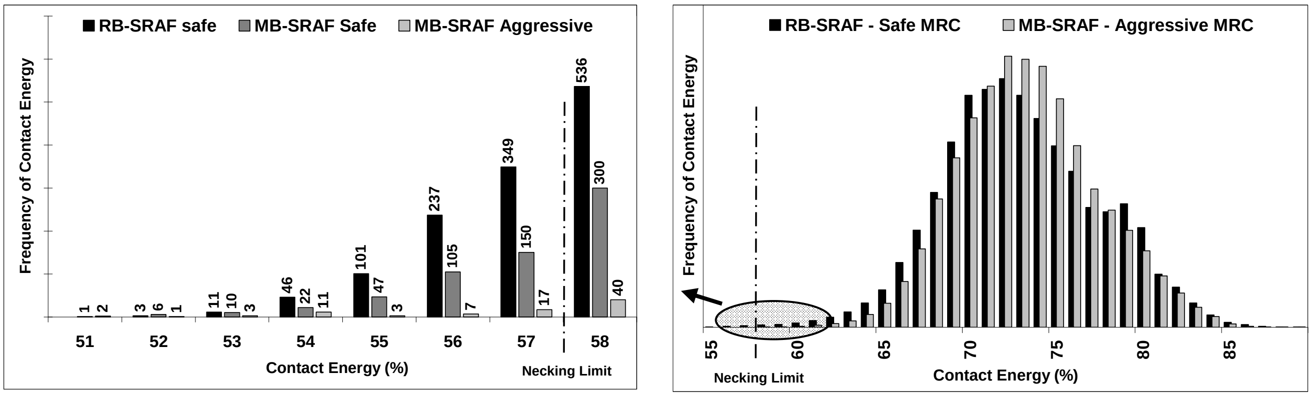


Figure 5. SRAF insertion rules

This model based approach has been applied on a real block, which is a typical test chip block used for OPC improvement. This is a high density design of around 19 mm^2. Two sets of SRAF MRC (Mask Rule Check) rules are presented and compared in this paragraph. A first trial is done with a MB-SRAF MRC rules set the same as with the one used for RB-SRAF, and a second trial is performed using more ‘aggressive’ MRC rules.

Due to the high development cost of creating a new SRAF rules set, the ‘aggressive’ RB-SRAF comparison was not performed.



a) Zoom of full chip histogram

b) Entire histogram

Figure 6. Contact Energy histograms of RB versus MB SRAF

The results of this comparison are shown in contact energy histograms. Figure 6.a is a zoom of the entire distribution close to the necking limit for the full chip rule based run (labeled “RB-SRAF safe”), the full chip model based run with strictly comparable rules (labeled “MB-SRAF Safe”), and the full chip model based run with aggressive rules (labeled “MB-SRAF Aggressive”). On figure 6.b, only ‘RB-SRAF Safe’ and ‘MB-SRAF Aggressive’ distribution are displayed for reading convenience: Safe RB and Safe MB full range histograms are very similar

With the same MRC rules, the model based approach is able to correct half of the defects indexed between 54% and 58% of contact energy. For most critical cases (below 54%) the enhancement is not obvious and has limitations, which helps to demonstrate the quality of the rule based approach. Nevertheless, a more aggressive model based approach leads to an almost defect free situation (see figure 6.a). The remaining defects with model based approach are attributed to global constraints in the generation of SRAF that could be enhanced with a local correction approach.

As a conclusion, we can say that MB-SRAF using aggressive MRC values seems to be a good candidate for testing the Tachyon FMO capability in resolving local HotSpots as an alternative to using MB-SRAF full chip, which has characteristics gathered in table 1.

Table 1. Advantages and disadvantages of MB-SRAF

Positive	Negative
allows correction of marginal hotspots	not efficient for all cases
decreases the total number of contact hotspots	requires powerful hardware or longer run times

4.2 RB-SRAF enhancement using MB-SRAF with Tachyon FMO

In previous section, we demonstrated that the full chip MB-SRAF engine may overcome a lot of hotspot configurations, but has also some limitations. A new flow is proposed in this paper, in order to cumulate the benefits of MB-SRAF placement optimization (around the hotspots generated by the RB-SRAF approach), the high quality boundary healing of FMO engine between the different correction areas, and the run time consideration by applying correction only in a small area around the hotspots. Our Flexible Mask Optimization flow is shown in table 2. This FMO flow has been applied on the same test case as the full chip test described in previous section.

Table 2. FMO flow (* LMC = Litho Manufacturing Check – verification)

1st Stage	**Standard OPC**	
	baseline OPC	initial RB-SRAF OPC
	baseline LMC (*)	extract list of initial hotspots
	HSFilter1	select hotspot list to enhance
2nd Stage	**1st 'aggressive' enhancement pass**	
	repairOPC1	enhancement using 'Aggressive' MB-SRAF
	repairLMC1	extract list of remaining hotspots
	HSFilter2	decide if a single pass is enough
3rd Stage	**2nd 'aggressive' enhancement pass**	
	repairOPC2	enhancement using 'Aggressive' MB-SRAF
	repairLMC2	extract list of remaining hotspots
	na	

Figure 7 shows the results in terms of hotspots number, i.e. contact energy below 58%. We can clearly see that the first pass of the flow is able to correct a large number of hotspots, shifting the worst cases on the right, whereas the second pass eliminates the remaining hotspots under the 58% limit. As a consequence, only one hotspot with 57% of contact energy is being seen after the entire flow, meaning that no new hotspot has been introduced during the correction, and that the enhancement boundary is perfectly managed by the engine. Focusing only on the hotspots correction, FMO leads to better results in terms of total defects number of the entire block than a full MB-SRAF (see table 3). This is due to the fact that a full chip application can result in conflicts from surrounding MB-SRAF generation whereas a highly localized correction with boundary healing can overcome these limitations.

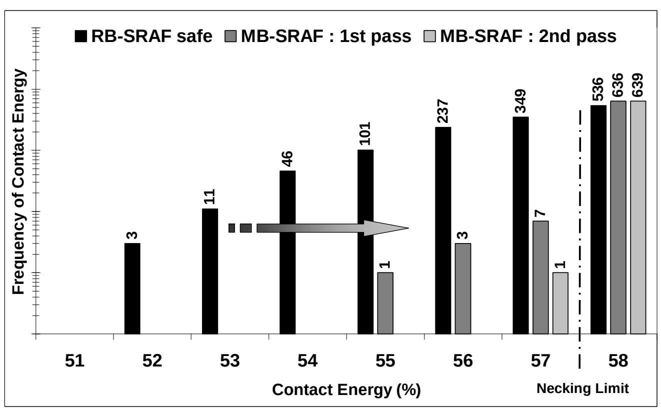


Figure 7. RB-SRAF Enhancement with MB-SRAF

Figure 8 shows the Exposure Latitude vs. Depth of Focus of the worst hotspot generated by the RB-SRAF initial solution (black line). For this particular example, using the FMO solution (with the 'aggressive' MRC rule set) enhances the process windows by 15%.

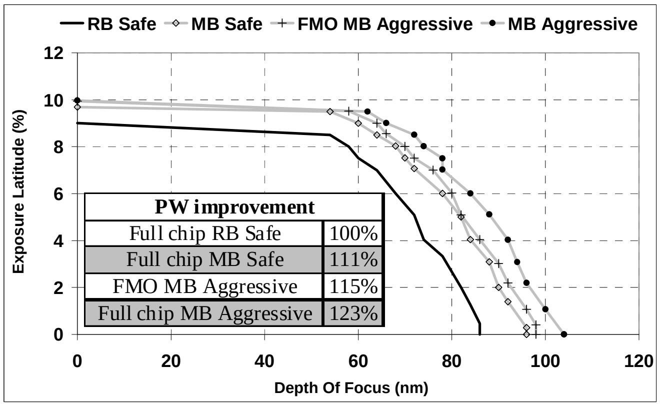


Figure 8. Full-chip PW results compared to FMO results

This FMO approach, i.e. using MB-SRAF to enhance the initial RB-SRAF worst hotspots, leads to a good compromise between run time spent to enhance worst hotspots and final defect values (target > 57%) (See table 3). Two MB-SRAF passes were necessary to obtain such results. Manipulation of hotspot is critical to obtain good final results, which is handled automatically by the "HSFilter" function built in to FMO. During the "RepairOPC" stage two key things occur: the advanced SRAF and OPC functions are performed on the hotspots, and secondly the boundary healing between the hotspot and the surrounding design area is performed. As a prerequisite one needs to define the type of HotSpots and the corresponding advanced OPC technique to be used to enhance them. Nevertheless, there is a compromise to find between better correction of the worst hotspot and loss of time, or vice-versa. The user may then decide or not to stop the FMO flow directly at the first pass if necessary.

Table 3. Run Time & Defect Summary

sum-up	worst hotspot	defect ratio	run Time
Full Chip RB-SRAF	**52.0%**	100.0%	**100%**
Full Chip Safe MB-SRAF	50.5%	46%	159%
Full Chip Aggressive MB-SRAF	50.8%	5.61%	173%
FMO MB-SRAF (1 pass)	55.0%	1.47%	127%
FMO MB-SRAF (2 pass)	**57.0%**	0.134%	**143%**

5. CONCLUSION

In this paper, STMicroelectronics and ASML Brion Technologies propose a new methodology of RB-SRAF hotspots enhancement using MB-SRAF. We have shown that a full model based approach is time consuming and strongly influenced by MRC parameters in order to get systematic enhancement for all cases. With the new FMO approach, the efficiency is improved by focusing on hotspots only, and we demonstrated this technique using a critical case for 2x nm node. We successfully illustrated that almost all (99.86%) of the hotspots can be enhanced (see table 3). It was also demonstrated that no new hotspots were introduced in the boundary regions where the different methods overlapped, which is very critical when choosing multiple correction methods on the same design.

The FMO flow is highly flexible and allowed us to find a perfect compromise between hotspots enhancement, efficiency and run time. FMO provides the option of an iterative approach. This is an entirely automatic flow that minimizes the manual analysis.

Based on the efficiency and the simplicity of the product, FMO is a good candidate for mask optimization improvement on other layers.

6. ACKNOWLEDGMENTS

The authors would like to thank Jean Noël Pena from STMicroelectronics for his contribution on OPC modeling.

REFERENCE

1. M.C. Tsai, S. Hsu, L. Chen, Y.W. Lu, J. Li, F. Chen, H. Chen, J. Tao, B.D. Chen, H. Feng, W. Wong, W. Yuan, X. Li, Z. Li, L. Li, R. Dover, H.Y. Liu, and J. Koonmen, "Full-chip source and mask optimization", Proc. SPIE 7973, 79730A (2011)
2. K. Hashimoto, S. Usui, S. Nojima, T. Kotani, E. Yamanaka, S. Tanaka, and S. Inoue, "Hotspot management and its applications in ultralow k1 lithography", J. Micro/Nanolith. MEMS MOEMS 8, 033007 (2009)
3. V. Farys, F. Chaoui, J. Entradas, F. Robert, O. Toublan and Y. Trouiller , "SRAF Enhancement using Inverse Lithography for 32 nm Hole Patterning and Beyond", Proc. SPIE 7488, (2009)